

Automated Classification of chest X-ray images as normal or abnormal using Convolutional Neural Network

Aayushi Gupta, Meddiff, Anupama C, Meddiff, P Indumathi, Meddiff, Anuj Kumar, Meddiff
anujkumar@icloud.com

Abstract— Chest X-Rays are generally used for diagnosing abnormalities in the thoracic area. Radiologists need to spend significant amount of time for interpreting scans. Automatic classification of these images could greatly help radiology interpretation process by enhancing real world diagnosis of problems. Hence, radiologists can focus on detecting abnormalities from the abnormal images rather than checking for it in all the images. In this paper, we present a machine learning approach to solve this problem. Here, the algorithm uses Convolutional Neural Networks (CNN) to learn and classify chest X-ray images as normal or abnormal based on image features.

Keywords—Classification, Radiology, Machine Learning, Convolution Neural Network, Chest X-Ray images.

I. INTRODUCTION

Thoracic diseases are one of the most common diseases in the world. According to the statistics of World Health Organization, about 1.6 million people die from lung cancer every year and 1.5 million die from lung tuberculosis.

A chest X-ray produces an image of the chest, lung, heart, ribs, airways and blood vessels. From chest X-ray image, trained radiologist can diagnose conditions such as consolidation, pneumonia, cardiomegaly, hiatal hernia, COPD, rib fracture, and so on [8]. Interpretation of chest X-rays heavily depend on the experience of radiologists since the diseased tissues in the images may be hidden by different body parts that overlap with one another, or lesions are in low contrast and overlap with large pulmonary vessels. These problems make chest X-ray images very difficult to read.

In radiology department, once image acquisition is done, it takes around 20 to 40 min to process, create evidence and interpret the images, and around 15 min to type and generate the report. Additional time is taken to review the reports and make changes, if necessary. But report distribution to the patient takes at least 24 hours after the image acquisition is done, due to the large volume of cases that radiologists have to go through. Since each chest X-ray takes a trained radiologist several minutes to review and write the report, and many radiologists have to work over-time, the chances of misdiagnosis due to exhaustion increases.

Traditional CADs are based on hand-crafted image features, and these features are then used to learn a binary or discrete classifier. The performance of such methods heavily depends on the extracted features, and it takes a long time for practitioners to come up with a good set of features, especially for complicated images like the X-rays.

Integrating machine learning techniques can significantly reduce radiology turnaround time by automating and making processing and interpretation of images easier and faster. Deep learning has been very successful for big data in the last few years, in particular for temporally and spatially structured data such as images and videos. Recent advances in storage, processing and computational power along with availability of large volumes of data have made deep learning very popular.

Convolutional Neural Network (CNN) is good at image analysis as they are designed to process 2D inputs. Being a type of deep learning technique, CNNs learn to detect spatial patterns in the training data and use it to make predictions on testing data.

In this paper, chest X-ray images from NIH dataset with labels were used to train a CNN model in order to classify them as either normal or abnormal.

II. METHODS

A. Dataset

A training dataset of 1942 chest X-ray images was collected from NIH website for this project. NIH website has a large base of X-ray, MRI and CT scan images along with corresponding reports publically available. Based on the findings in the reports, the images were segregated as normal or abnormal. A total of 1000 abnormal and 942 normal images were used. The abnormalities present in this dataset includes consolidation, atelectasis, pulmonary markings, lung nodules, cardiomegaly, calcified granuloma, emphysema, pleural effusion and rib fracture as shown in “Fig. 1.”. The dataset was split as 1642 for training and 300 for validation of the model. Further, the model was tested with unlabeled images to check the model’s performance.

B. Model Architecture

Here, a CNN model is used to classify chest X-ray images as normal or abnormal.

Typically, a CNN [1] [2] consists of three types of layers that are stacked on top of each other. They are convolution, max pooling and fully connected/dense layers. Convolution layer consists of a set of learnable filters (or kernels) that convolves through the input volume to extract features. Max pooling layer performs non-linear down-sampling of images where pixels are divided into groups and then one with the largest value is retained. Fully connected layer gives a high level reasoning, where every neuron of this layer is connected to every neuron in the previous layer.

The architecture of this model is as shown in "Fig. 2," and consists of six convolution layers and two fully connected layers with Rectified Linear Unit (Relu) as activation function. A max pooling layer follows each convolution layer to down-sample the images. A dropout layer is added in order to deactivate certain neurons that minimizes overfitting of the model. The last fully connected layer contains softmax activation in order to get probability of an image belonging to a particular class.

Since CNN learns by supervised learning, a regression layer minimizes loss of the model using Adam optimizer.

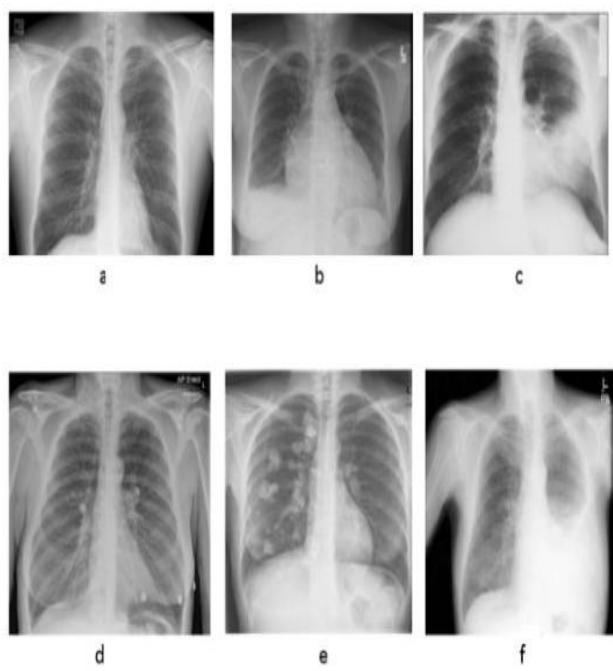


Fig.1. Figure showing sample abnormal chest images present in the training dataset. a) Pulmonary markings, b) Cardiomegaly, c) Lung consolidation, d) Calcified granuloma, e) Lung nodules, f) Pleural effusion.



Fig.2. Architecture of proposed model.

C. Training

The CNN model was trained with 1642 chest X-ray images and the remaining 300 images were used to validate the performance of the model after it has been trained. Since CNN learns by supervised training, labels associated with each image were given in order to classify them. The input images were loaded, resized and fed into the network in batches of 64. Then the model was trained using backpropagation [3], where the output obtained from the model was compared with the desired output (labels in the form of one hot array) to get loss. Adam optimizer was used to minimize this loss by updating weights in the network. Training accuracy and loss was obtained for each epoch.

After each of the training iterations, the model performance was evaluated with validation set to get validation accuracy and loss.

Thus the model, after training, can detect patterns present in unseen chest X-ray images.

D. Testing

The trained model is tested with 100 normal and 100 abnormal images, which were not part of the training dataset, to get confusion matrix for the predictions given by the model. From confusion matrix, performance metrics like sensitivity, specificity and testing accuracy were calculated.

E. Observations

The output of convolution layer of CNN shows the features in the image that were highlighted for a particular filter, called activation maps. These activation maps of different convolution layers from the model are as shown in “Fig. 3.”

F. Evaluation Metrics

We used the following metrics to evaluate the performance of the model.

1. *Confusion Matrix* [9] (as shown in “Table I”) contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using data in the matrix. The following table shows the confusion matrix for a two class classifier.

The entries in the confusion matrix have the following meaning:

- True Positive (TP) is the number of correct predictions that an instance is normal,
- False Negative (FN) is the number of incorrect predictions that an instance is abnormal,
- False Positive (FP) is the number of incorrect predictions that an instance is normal,
- True Negative (TN) is the number of correct predictions that an instance is abnormal.

2. *Accuracy (AC)* is how close the measured value is to the actual (true) value, i.e., the proportion of the total number of predictions (classifications) that were correct. It is determined using the equation:

$$AC = (TP+TN) / (TP+TN+FP+FN)$$

3. *Sensitivity (True Positive Rate, SN)* is the ability of a test to correctly identify those with abnormality, i.e., how well the system has correctly predicted abnormal cases.

$$SN = TP / (TP+FN)$$

4. *Specificity (True Negative Rate, SP)* is the ability of the test to correctly identify those without the abnormality (normal cases), i.e., how well the system has correctly predicted normal cases.

$$SP = TN / (TN+FP)$$

	Predicted Normal (Positive)	Predicted Abnormal (Negative)
Actual Normal (Positive)	TP	FN
Actual Abnormal (Negative)	FP	TN

Table showing confusion matrix

III. RESULTS

The model was trained for 10, 20, 30, 40 and 50 epochs. The training and validation accuracies for all of these are as shown in “Table II”. It was seen that model trained for 50 epochs gave an overall better result than the model trained for other epochs.

The model trained for these epochs were tested with 100 normal and 100 abnormal chest X-ray images to get metrics values: testing accuracy, sensitivity and specificity, as shown in “Table III”. Here, it was seen that model trained for 50 epochs gave a better testing accuracy than the model trained for other epochs.

TABLE II

Number of epochs	Training accuracy (%)	Training loss	Validation accuracy (%)	Validation loss
10	57.59	0.6649	51.00	0.7113
20	74.00	0.5324	61.67	0.6690
30	86.96	0.3697	55.67	0.0639
40	94.95	0.1545	59.00	1.552
50	96.61	0.0938	59.30	2.022

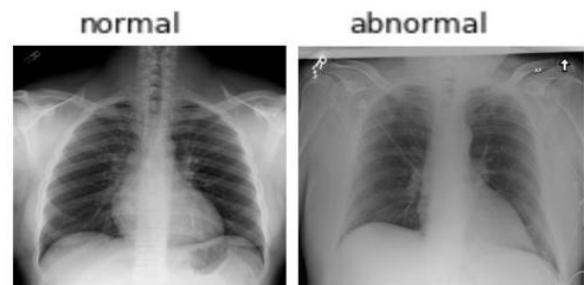
Table shows model training output with respect to number of epochs (Input parameter of the model)

TABLE III

Number of epochs	Testing accuracy (%)	Sensitivity	Specificity
10	51.50	0.12	0.91
20	56.00	0.68	0.44
30	58.00	0.65	0.51
40	52.00	0.33	0.72
50	61.00	0.74	0.48

Table shows comparison of testing metrics with respect to number of epochs of the trained model

Fig.3. Activation map of an image: A plot showing the features extracted in the convolution layer 1.



REFERENCES

- [1] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *papers.nips:4824*, December 03 - 06, 2012.
- [2] Jianxin Wu, "Introduction to Convolutional Neural Networks," National Key Lab for Novel Software Technology Nanjing University, May 1, 2017 *cs.nju.wujx*.
- [3] Dr. Rama Kishore,Taranjit Kaur, "Backpropagation Algorithm: An Artificial Neural Network Approach for Pattern Recognition," *International Journal of Scientific & Engineering Research, Volume 3, Issue 6, June2012: ISSN 2229-5518*.
- [4] Ms. Sonali. B. Maind and Ms. Priyanka Wankar, "Research Paper on Basic of Artificial Neural Network," *International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2 Issue: 1, January 2014, ISSN: 2321-8169*.
- [5] Saeed AL-Mansoori, "Intelligent Handwritten Digit Recognition using Artificial Neural Network," *Int. Journal of Engineering Research and Applications, ISSN : 2248-9622, Vol. 5, Issue 5, (Part -3) May 2015, pp.46-51*.
- [6] Chen Wang and Yang Xi, "Convolutional Neural Network for Image Classification," Johns Hopkins University Baltimore, MD 21218, *cs.jhu: ~cwang107*.
- [7] Deepika Jaswal and Sowmya.V, K.P.Soman, "Image Classification Using Convolutional Neural Networks," *International Journal of Scientific & Engineering Research, Volume 5, Issue 6, June-2014 ISSN: 2229-5518*.
- [8] Yuxi Dong, Yuchao Pan, Jun Zhang and Wei Xu, "Learning to Read Chest X-Ray Images from 16000+ Examples Using CNN," *Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2017 IEEE/ACM International Conference, July 2017, DOI: 10.1109/CHASE.2017.59*
- [9] A. K. Santra, C. Josephine Christy, "Genetic Algorithm and Confusion Matrix for Document Clustering," *IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 1, No 2, January 2012*.

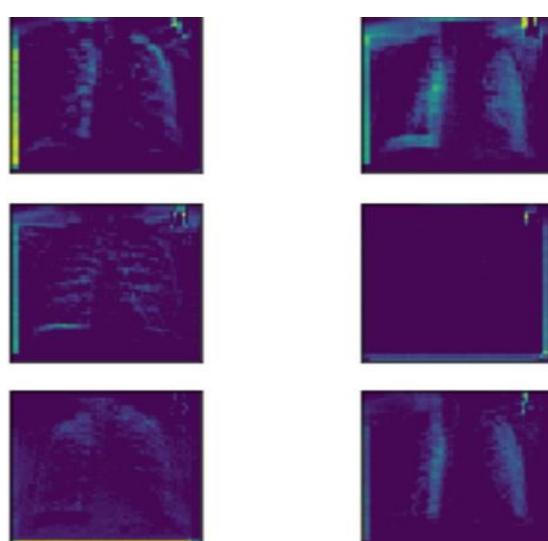


Fig.4.Result: Test image outputs as class labels as predicted by the trained model.